



Codificación y Decodificación en Poblaciones Neuronales.

- Es razonable pensar que la información sobre el mundo externo se represente o se procese de una forma distribuida en toda la completitud del cerebro.
- Anteriormente hemos considerado varios esquemas de codificación.
- Aquí vamos a considerar la distribución espacial de la información en una población de neuronas.
- Utilizaremos un marco probabilístico debido a la mencionada variabilidad neuronal de los sistemas biológicos.



Codificación y Decodificación en Poblaciones Neuronales.

- **Codificación:** Es la forma específica en la que el estímulo es representado en la neurona o población neuronal.
- **Decodificación:** Es la manera en que el mensaje que está distribuido en la población puede ser interpretado de acuerdo a la respuesta de la neurona o población neuronal.



Codificación en Poblaciones Neuronales.

- Podemos considerar la codificación de un patrón de estímulos en función de una probabilidad de respuesta de la población neuronal.

$$P(r|s)=P(r_1^s, r_2^s, r_3^s,|s),$$

donde s es cierto estímulo y r_i^s es la respuesta específica de la neurona i respecto al estímulo s .

- Principalmente, se consideran para las respuestas frecuencias o tasas de disparos, ya que son fáciles de medir.
- Sin embargo, otro tipo de medida para la respuesta neuronal es completamente válido.
- P puede ser una distribución de probabilidad continua o discreta.



Decodificación Bayesiana en Poblaciones Neuronales.

- La decodificación es la inversa a la codificación, es decir queremos saber qué estímulo fue presentado a la población neuronal sabiendo que tenemos una respuesta neuronal determinada.
- Esto puede ser expresado de la siguiente forma:
$$P(\mathbf{s}|\mathbf{r})=P(\mathbf{s}|\mathbf{r}_1^s, \mathbf{r}_2^s, \mathbf{r}_3^s, \dots).$$
- Por tanto, si conocemos esta distribución de probabilidad, parece lógico que digamos que el **estímulo que corresponda al máximo** de esa distribución de probabilidad, sea precisamente el estímulo que estamos buscando:
$$\underline{\mathbf{s}}=\max_{\mathbf{s}}(P(\mathbf{s}|\mathbf{r}))$$

Decodificación Bayesiana en Poblaciones Neuronales.

- Pero aquí tenemos un problema: ¿**Cómo estimamos $P(s|r)$** ? Estimar $P(r|s)$ de los registros neuronales experimentales es relativamente “fácil”.
- El teorema de **Bayes** relaciona $P(s|r)$ con $P(r|s)$
$$P(s|r) = (P(r|s)P(s))/P(r).$$
- Así, además de $P(r|s)$ necesitamos medir las distribuciones $P(s)$ y $P(r)$.
- Esto se puede hacer, pero no es trivial y resulta muy costoso.
- Existe un “truco” (***Maximum Likelihood Estimate***) si $P(s)$ es constante, que consiste en: maximizar respecto a s la distribución $P(s|r)$ es lo mismo que maximizar $P(r|s)$ (debido a que el T. Bayes relaciona estas dos cantidades).



Decodificación Bayesiana en Poblaciones Neuronales.

- Así, cuando $P(s)$ es constante

$$\underline{s}_{ML} = \max_s (P(r|s)) ,$$

viendo $P(r|s)$ como una función de s cuando tenemos una respuesta fija (es decir una muestra representativa de la respuesta).

- Es decir buscamos el estímulo que hace más “verosímil” la muestra representativa de la respuesta.
- Vamos a recordar un poco que es la verosimilitud.
- Así hacemos un paréntesis, con un poquito de estadística.



Estimadores de máxima verosimilitud: distribución conjunta de la muestra

- Los conceptos de funciones de verosimilitud se deben a Fisher, y es fundamental en inferencia estadística.
- Este concepto se define a partir de la distribución conjunta de la muestra.
- Supongamos una variable discreta x con distribución $P(x; \vartheta)$ que es conocida.
- ϑ son los posibles parámetros de distribución (por ejemplo media y varianza en una distribución normal).
- Supongamos que tomamos muestras independientes de tamaño n , representando está por el vector \mathbf{X} .

Estimadores de máxima verosimilitud: distribución conjunta de la muestra

- Así podemos definir la distribución conjunta de la muestra en función de esta variable, y para el caso de una muestra aleatoria simple tenemos:
 - $P(\mathbf{X} = \mathbf{X}_0) = P(x_1 = x_{10}, x_2 = x_{20}, \dots, x_n = x_{n0}) =$
 $= P(x_{10}) \times \dots \times P(x_{n0})$
 - Una muestra es aleatoria simple (m.a.s.), si :
 - Cada elemento de la población tiene la misma probabilidad de ser elegido.
 - Las observaciones se realizan con reemplazamiento (población idéntica en todas las extracciones).
 - La primera condición asegura representatividad de la muestra (si A esta en el 20% y todos los elementos tienen idéntica probabilidad de ser seleccionados, la muestra tendrá un 20% también de A).
 - La segunda se impone por simplicidad.
- Por tanto, conociendo la distribución $P(x; \vartheta)$ podemos calcular fácilmente la probabilidad de cualquier muestra.



Estimadores de máxima verosimilitud: distribución conjunta de la muestra

- En el caso continuo (función de densidad $f(x; \vartheta)$), la probabilidad del intervalo $x_i - 1/2, x_i + 1/2$, la podemos aproximar por el rectángulo de altura $f(x_i)$ y base unidad:
- $P(x_i) = f(x_i) \cdot 1$
- Por tanto la probabilidad de la muestra aleatoria simple:
- $P(x_1, \dots, x_n) = \prod f(x_i)$
- Así la función de densidad conjunta de la muestra $f(x_1, \dots, x_n)$ se interpreta como la probabilidad de obtener los valores muestrales $x_1 \pm 0,5, \dots, x_n \pm 0,5$.



Estimadores de máxima verosimilitud: la función de verosimilitud

- Sea una variable aleatoria continua x con función de densidad que $f(x|\vartheta)$ para indicar que depende de un vector de parámetros ϑ . Es decir dado que conozco ϑ , representa cual es la función de densidad de la variable aleatoria x .
- Si tenemos una muestra aleatoria simple $X = (x_1, \dots, x_n)$, entonces la función de densidad conjunta de la muestra es:
- $f(X|\vartheta) = \prod f(x_i|\vartheta)$
- Es decir cuando conozco ϑ la expresión anterior determina la probabilidad de aparición de cada muestra.



Estimadores de máxima verosimilitud: la función de verosimilitud

- En inferencia para un problema de estimación se conoce un valor particular de una muestra X , siendo desconocido el parámetro ϑ .
- Así si sustituimos X por el valor observado de una muestra, $X_0 = (x_{10}, \dots, x_{n0})$, entonces la función $f(X_0|\vartheta)$ puede ser vista como una función del parámetro.
- Observar que $X_0 = (x_{10}, \dots, x_{n0})$, es la muestra, y no varia una vez que la fijamos.
- Es decir $f(X_0|\vartheta)$ puede ser visto y proporciona, para cada valor de ϑ , la probabilidad de obtener el valor muestral X_0 para ese ϑ .

Estimadores de máxima verosimilitud: la función de verosimilitud

- Así esta nueva función que obtenemos cuando variamos ϑ , mientras mantenemos X_0 fijo (no variamos la muestra), define la **función de verosimilitud**, $\ell(\vartheta|X)$ (dado que conozco la muestra como varia la probabilidad en función del parámetro ϑ).
- $\ell(\vartheta|X)$, o $\ell(\vartheta)$: Es decir $\ell(\vartheta|X)=\ell(\vartheta)=f(X_0|\vartheta)$, con X_0 fijo y ϑ variable.
- La óptica del problema cambia, en vez de tener un parámetro fijo ϑ y calcular para ese parámetro la probabilidad de obtener distintas muestras X , lo que fijamos es una determinada muestra X_0 y estimamos que valor del parámetro ϑ hace más verosímil la muestra que se observa X_0 .



Estimadores de máxima verosimilitud: la función de verosimilitud

- Este enfoque cambia completamente la forma de la función.
- Si tenemos una variable x que distribuye según una Poisson:
 - $P(x = r) = \frac{\lambda^r}{r!} e^{-\lambda}, r = 0, 1, 2, \dots$, y observamos el valor de la muestra $x=5$, entonces $\ell(\lambda) = \frac{\lambda^5}{5!} e^{-\lambda}$, es la función de verosimilitud para una muestra de un solo valor de $x=5$.



Estimadores de máxima verosimilitud: la función de verosimilitud

- Esta función $\ell(\lambda)$ es continua en λ y proporcional a la probabilidad de observar $x=5$ para cada valor posible de λ .
- El valor de la verosimilitud no es único:
 $\ell(\vartheta_1) = f(X_0 \mid \vartheta_1) > f(X_0 \mid \vartheta_2) = \ell(\vartheta_2)$.
- Esto quiere decir que a la vista de los datos muestrales el valor del parámetro ϑ_1 es más verosímil que el valor del parámetro ϑ_2 ya que la probabilidad de obtener la muestra observada X_0 es mayor con ϑ_1 que con ϑ_2 .



Estimadores de máxima verosimilitud: la función de verosimilitud

- Observar que la verosimilitud tiene unidades, las de la variable x , entonces la diferencia de verosimilitudes no tiene sentido ya que varía arbitrariamente en función de las unidades de la variable x .
- Por lo tanto para comparar verosimilitudes lo mejor es el cociente de las mismas ya que este es invariante frente a las diferentes unidades de la variable:
- $\ell(\vartheta_1|X)/\ell(\vartheta_2|X)$, este cociente es invariante hacia cambios de escalas en la variable x que se está observando.
- El cociente $\ell(\vartheta_1)/\ell(\vartheta_2)$ se puede sustituir por la diferencia de logaritmos: $\ln\ell(\vartheta_1) - \ln\ell(\vartheta_2)$
- Así se puede definir la **función soporte** por el logaritmo de la verosimilitud: $L(\vartheta) = \ln\ell(\vartheta)$.



Estimadores de máxima verosimilitud: la función de verosimilitud

- Resumiendo:
 - La función de verosimilitud es la herramienta básica que nos permite juzgar la compatibilidad entre los valores muestrales observados y los posibles valores del parámetro de la distribución de probabilidad.
 - Si queremos comparar dos posibles valores del parámetro, ϑ , se debe utilizar el cociente de sus verosimilitudes, y no su diferencia, ya que la diferencia depende de la escala de medida de las variables, como hemos dicho antes.

Estimadores de máxima verosimilitud: la función de verosimilitud

- Por ejemplo para estimar el parámetro, λ , de Poisson de la muestra observada x_1, \dots, x_n hacemos:
 - $P(x) = \frac{\lambda^x}{x!} e^{-\lambda}$, distribución de Poisson.
 - $\ell(\lambda) = \prod P(x_i|\lambda) = P(x_1|\lambda)P(x_2|\lambda) \dots P(x_n|\lambda) = \frac{\lambda^{\sum x_i}}{\prod x_i!} e^{-n\lambda}$.
- Como el término $1/(\prod x_i!)$ es una constante la podemos eliminar y escribir la función de verosimilitud para la distribución de Poisson como:
 - $\ell(\lambda) = e^{-n\lambda} \lambda^{\sum x_i} = e^{-n\lambda} \lambda^{n\bar{x}}$.
 - Así la función soporte será para la distribución de Poisson vendrá dada por: $L(\lambda) = -n\lambda + n\bar{x} \ln \lambda$.

Estimadores de máxima verosimilitud: la función de verosimilitud

- Una vez que tenemos calculada una función de verosimilitud para un vector de parámetros $\boldsymbol{\vartheta}$, $\ell(\boldsymbol{\vartheta})$, un procedimiento intuitivo para estimar los parámetros de la distribución a partir de los valores observados muestralmente es maximizar el valor del parámetro que sea más verosímil, es decir el que maximice la verosimilitud.
- Así podemos resolver el sistema de ecuaciones:
 - $\partial \ell(\boldsymbol{\vartheta}) / \partial \vartheta_1 = 0, \dots, \partial \ell(\boldsymbol{\vartheta}) / \partial \vartheta_p = 0$
- El valor que resuelve el sistema de ecuaciones, $\hat{\boldsymbol{\vartheta}}$, corresponderá a un máximo si el valor de la matriz hessiana de segundas derivadas en ese punto es definida negativa: $H(\hat{\boldsymbol{\vartheta}}) = (\partial^2 \ell(\boldsymbol{\vartheta}) / \partial \vartheta_i \partial \vartheta_j)_{\boldsymbol{\vartheta} = \hat{\boldsymbol{\vartheta}}}$.



Decodificación mediante Curvas de Ajuste en Poblaciones Neuronales.

- Cuando no podemos ni calcular $P(r|s)$ de una forma fácil y eficiente, ya no podemos calcular ni el *coding* ni el *decoding* en una población neuronal.
- Pero recordemos las **curvas de ajuste** que ya habíamos visto al principio del tema.
- Recordemos que estas curvas nos relacionan para una neurona el estímulo y su respuesta, de la siguiente manera: $r_i = f_i(s)$.
- Además, estas curvas se pueden medir de una manera flexible y razonable para muchas neuronas.
- Normalmente, estas curvas se suelen ajustar muy bien mediante gaussianas.



Decodificación mediante Curvas de Ajuste en Poblaciones Neuronales.

- Aunque, de nuevo, tenemos un problema. Para decodificar (para un valor determinado de *firing rate*), tenemos **ambigüedad** (cierto *firing rate* puede ser resultado de 2 estímulos).
- Cuando contemplamos la **población** esa ambigüedad desaparece.
- Para ello, necesitamos una segunda neurona con la curva de ajuste ligeramente desplazada y medir su *firing rate*.

Decodificación mediante Curvas de Ajuste en Poblaciones Neuronales.

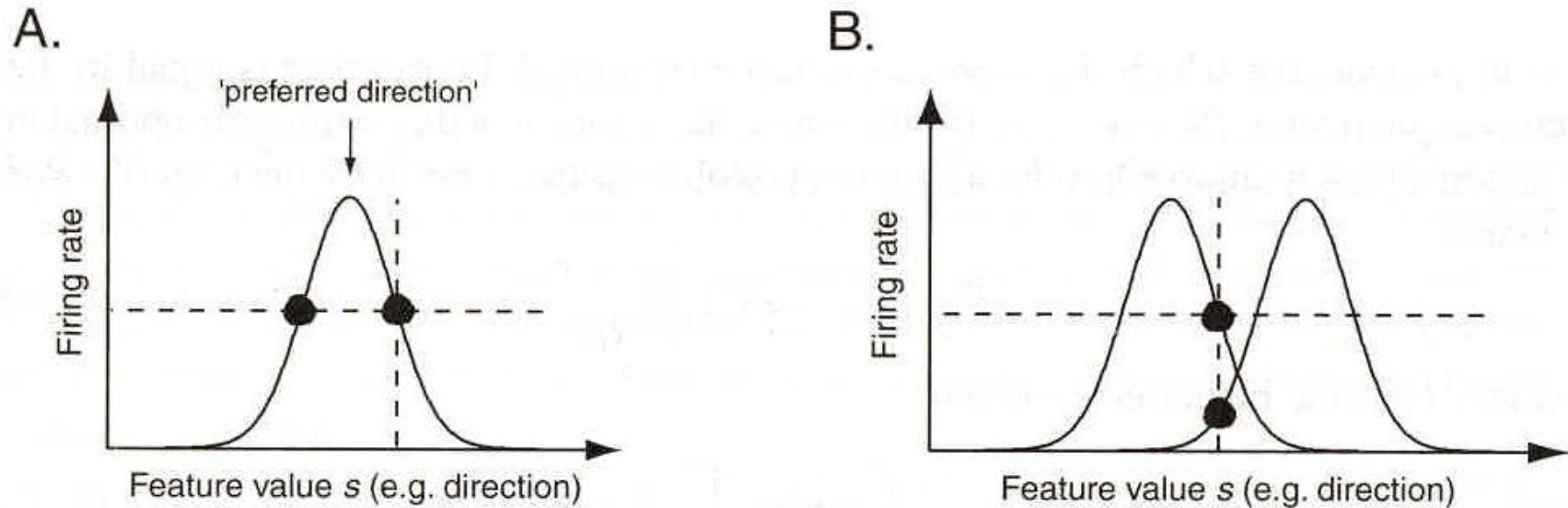


Fig. 5.13 Gaussian tuning curves representing the firing rate of a neuron as a function of a stimulus feature. (A) A single neuron cannot unambiguously decode the stimulus feature from the firing rate. (B) A second neuron with shifted tuning curve can resolve the ambiguity.



Decodificación mediante Curvas de Ajuste en Poblaciones Neuronales.

- Observemos que la **precisión de la decodificación depende de los valores del estímulo**.
- El cálculo de la decodificación tendrá mayor resolución y fiabilidad cuando tengamos estímulos en la **máxima variación de las pendientes de las curvas**.
- Hemos visto diversas técnicas para estimar la decodificación. Hay muchas más, y la utilización de una u otra dependerá del problema exacto que estemos atacando en cada momento.



Representación Distribuida de Información en el Cerebro

- En primer lugar, vamos a ver cómo se puede distribuir un estímulo en el cerebro potencialmente.
- Supongamos que **representamos el estímulo mediante un vector, siendo las componentes del mismo valores que miden las respuestas de diferentes áreas del cerebro** (por ejemplo, *firing rate*).
- Queremos saber cuántas componentes están activas y por tanto son usadas para representar un estímulo en el cerebro.
- Vamos a llamar a cada componente un nodo o “cluster” del cerebro.
- ¿Qué tipo de estrategias podrían existir?
Representación **local**, representación totalmente **distribuida** y un compromiso entre las otras dos, es decir, una representación **esparcida**.



Representación Distribuida de Información en el Cerebro

- **Representación local:** cuando se presenta un estímulo particular, solamente un nodo o “cluster” representa un estímulo, así sabiendo qué nodo (o neurona) está activado sabemos el estímulo que se presenta. Las neuronas con estas características se han denominado en la literatura ***grandmother cells***.
- Un vector de longitud N podría representar N diferentes características con tal representación.
- Así el **número de estímulos** que pueden ser codificados con este tipo de representación, **escala linealmente con la dimensión del vector**.
- En el caso de representaciones con nodos binarios “on” o “off”, solamente un nodo podría estar a “on” para cada posible estímulo.



Representación Distribuida de Información en el Cerebro

- **Representación totalmente distribuida:** Este tipo de representación puede verse como el extremo totalmente contrario. Aquí un estímulo es codificado mediante un combinación de todas las componentes del vector. El **número total de representaciones se escala exponencialmente con el número de nodos.**
- El número de representaciones de las codificaciones puede ser mucho mayor.
- En el caso binario, podríamos pensar que la probabilidad de cada nodo de estar en “on” u “off” es del 50%, así el número de nodos activos es del 50% en promedio.
- Una ventaja de esta representación es podemos definir **similitudes** entre estímulos, simplemente contando cuantas componentes del vector tienen valores similares. Esto es importante para construir asociaciones.



Representación Distribuida de Información en el Cerebro

- **Representación esparcida:** es un compromiso entre las dos anteriores representaciones. **Solamente una fracción de los nodos** se ve involucrada en la representación de un determinado estímulo.
- El número de estímulos que se pueden representar esta comprendido entre los dos casos de anteriores.
- En el caso de vectores binarios podríamos tener gran probabilidad de los nodos a “off” en comparación con los nodos a “on”.
- Queremos especificar más cuantitativamente cuantas neuronas están implicadas en el procesamiento neuronal de un estímulo individual.
- Para definiremos una nueva cantidad, o medida que denominaremos “**esparcimiento**”.
- Esta definición será más fácil y obvia para el caso de nodos binarios.



Representación Distribuida de Información en el Cerebro

- Supongamos nodos binarios (on-off), y supongamos que $r_i=1$ neurona i disparando, y $r_i=0$ neurona i en silencio. Así, una medida del esparcimiento en la representación sería:

$$a = (1/S) \sum_s (1/N) \sum_i r_i^s$$

Donde S es el conjunto de estímulos donde se evalúa el esparcimiento, N es número de neuronas que se considera en la población, y r_i^s la respuesta específica de la neurona i hacia el estímulo individual s .

- **Esta medida representa el número promedio de neuronas que están disparando a través de un conjunto de estímulos.**



Representación Distribuida de Información en el Cerebro

- En otras palabras, si ahora consideramos los correspondientes *firing rates* con $r=0$ si no responde la neurona, y $r=1$ si responde, entonces el promedio de la tasa de disparo relativa es $a = \langle r_i^s \rangle_{i,s}$.
- El promedio anterior se toma sobre el promedio de neuronas en la población y el número de estímulos en el conjunto de test.
- Una representación **completamente distribuida** tiene para esta definición de esparcimiento un valor de **$a=0,5$** .
- Esta medida es menor si nos restringimos a que el número de nodos que pueden estar a “on” es menor que $N/2$. Hay que notar que el otro caso de tener nodos a “on” mayor que $N/2$, es el mismo problema sin más que redefinir los valores para la representación “on” y “off”.

Representación Distribuida de Información en el Cerebro

- La definición del “esparcimiento” para el caso de vectores con **componentes continuas** no es obvio.
- Tenemos que decidir cuanto sopesamos las tasas de disparo grandes frente a las pequeñas.
- Lo que deberíamos de hacer es tomar información en la tasa de disparo **respecto una especie de varianza de la tasa de disparo**.
- Así la nueva medida la podríamos definir, promediada sobre un con conjunto de estímulos de nuevo, como:
$$a = \langle r_i^s \rangle_{i,s}^2 / \langle (r_i^s)^2 \rangle_{i,s}.$$
- Por ejemplo, podemos medir la tasa de disparo para N neuronas en respuesta a un número de estímulos S y estimar el esparcimiento como:
$$a = ((1/S) \sum_s (1/N) \sum_i r_i^s)^2 / ((1/S) \sum_s (1/N) \sum_i (r_i^s)^2)$$
- Esta expresión y la anterior (2 transparencias atrás), son equivalentes cuando estamos en el caso de nodos binarios con componentes de 0 y 1.



Representación Distribuida de Información en el Cerebro

- ¿Qué es una **característica**?
- Hemos hablado sobre la respuesta de una población de nodos a un estímulo.
- Tal estímulo podría ser una imagen como una cara.
- Cuando investigamos las respuestas neuronales a estímulos externos es obvio intentar descomponer el estímulo en diferentes características para encontrar a que característica esta respondiendo una neurona en concreto.
- Por ejemplo podríamos investigar como una neurona responde a caras con estímulos hechos de partes de la cara, tales como los ojos, la boca, y incluso combinaciones más complicadas de las diferentes características.



Representación Distribuida de Información en el Cerebro

- De hecho esto es lo que se hace en muchos experimentos neurofisiológicos.
- Si las neuronas son más específicas a ciertas características, entonces es obvio que solamente un conjunto de neuronas que respondan a una cara también responderán a un conjunto de características.
- Así podemos estudiar el número de nodos que responden a un estímulo solamente en función de un grano más fino de la naturaleza del estímulo.
- Una definición alternativa de característica podría ser entender la representación poblacional de un estímulo como una descomposición en características del estímulo.
- Cada componente del vector de la población podría definir una característica. De hecho llamaremos al vector de la población vector de características.
- En este contexto podríamos tener un estímulo que active solamente un número muy pequeño de células (un solo nodo en caso extremo).



Representación Distribuida de Información en el Cerebro

- Así por definición la representación podría ser local.
- Aunque sin embargo es muy probable que incluso un estímulo altamente especializado provocaría disparos de un número de células en el cerebro.
- Por ejemplo las células que tienen una representación mediante las curvas de ajuste (ver figuras anteriores), las cuales demuestran que responden de forma gradualmente a ciertas características, podrían tener solapamiento de las curvas de ajuste.
- Incluso una característica muy específica se distribuye entre neuronas, y las representaciones de los estímulos naturales con muchas características son probablemente basadas en un código distribuido en el cerebro.



Representación Distribuida de Información en el Cerebro

- De hecho muchos experimentos neuropsicológicos sobre las funciones cerebrales sugieren que la información esta representada en una forma distribuida en cerebro.
- Vamos a intentar en lo que sigue estimar como de grande es la típica representación distribuida en el cerebro.
- Es decir cuantas neuronas hay implicadas en un proceso particular del cerebro tal como la representación de un objeto visual en corteza temporal inferior.
- Este tipo de preguntas es bastante difícil responder, ya que es difícil el registro de muchas neuronas simultáneamente.



Representación Distribuida de Información en el Cerebro

- Una técnica que se utilizan los tintes en neuronas.
- Así registros ópticos con y sin tintes sensibles al voltaje muestran las áreas de actividad de las neuronas.
- Hay que tener en cuenta que esas áreas están compuestas por cientos de células que deberían ser registradas aisladamente para estudiar bien la distribución del código.
- Otro problema de los registros ópticos es que podía no mostrar pequeñas áreas, que participan activamente en la representación distribuida del objeto.



Representación Distribuida de Información en el Cerebro

- Aquí, en este tipo de problema, también podemos utilizar teoría de información.
- De hecho es una herramienta fundamental en este tema.
- Nos podemos preguntar cuanta información se gana sobre el estímulo externo a través de la actividad de grupo de neuronas.
- Variando el número de neuronas que entran en el cálculo de información se podría inferir el orden de la magnitud del número de neuronas involucradas en la representación del objeto en cuestión.



Representación Distribuida de Información en el Cerebro

- Para estudiar la naturaleza de las representaciones en el cerebro queremos estimar la información mutua entre las poblaciones de neuronas o entre un conjunto de estímulos externos y una población de neuronas.
- De nuevo tenemos la explosión del espacio de probabilidad para estimar la probabilidad conjunta $P(y) = P(r_1^s, r_2^s, r_3^s, \dots)$, respuesta de la población neuronal a un estímulo.
- Necesitamos estimar correctamente esta distribución de probabilidad para calcular $MI(s, y) = \sum_s \sum_y P(s, y) \log_2 (P(s, y) / (P(s)P(y)))$

Representación Distribuida de Información en el Cerebro

- En vez de evaluar esta información mutua, $MI(s,y)$, directamente, utilizamos otra estrategia.
- Podemos intentar **reconstruir la señal neuronal “y”** (son las respuestas de las neuronas de la población) de la señal original. De forma más precisa queremos calcular una estimación del estímulo \underline{s} (a través de los métodos anteriores de decodificación) $MI(s,\underline{s}) = \sum_s \sum_{\underline{s}} P(s,\underline{s}) \log_2 (P(s,\underline{s}) / (P(s)P(\underline{s})))$. Es decir dado que yo tengo una respuesta “y” cuál es el estímulo que maximiza esa respuesta: $\underline{s} = \max_s (P(s|y))$.
- Esta expresión, $MI(s,\underline{s})$, es equivalente a la anterior $MI(s,y)$, siempre que la decodificación se haga perfectamente. Es decir sin pérdida de información. Si hay pérdida de información esta expresión, $MI(s,\underline{s})$, es más baja que la anterior, $MI(s,y)$.



Representación Distribuida de Información en el Cerebro

- Para estimar el contenido de información en un conjunto de neuronas, uno tiene que grabar de varias neuronas en un área simultáneamente si las respuestas de las neuronas de la población no son independientes unas de otras.
- Si por otro lado son independientes podemos utilizar los registros para estimar directamente la información.
- Es lo que ha hecho Edmund Rolls y sus colegas en el IT cortex para estimar la información.
- Los autores usan neuronas que responden a 20 caras y utilizan un estimador estadístico para el proceso de decodificación ajustando a gaussianas las fluctuaciones de la respuesta neuronal de cada célula.



Representación Distribuida de Información en el Cerebro

- Se muestran los resultados entre la 20 caras como conjunto de estímulos en función de la respuesta de C neuronas registradas (símbolos estrella en la figura A).
- Vemos que la información crece rápidamente con el número de neuronas que se tienen en cuenta en el análisis de información.
- La información de 20 caras (puntos) puede ser comparada con la información de 8 caras (cruces) y de 4 caras (cuadrados).
- Así variando el número de neuronas podemos extrapolar cuál es el orden de magnitud de neuronas implicadas en el proceso.

Representación Distribuida de Información en el Cerebro

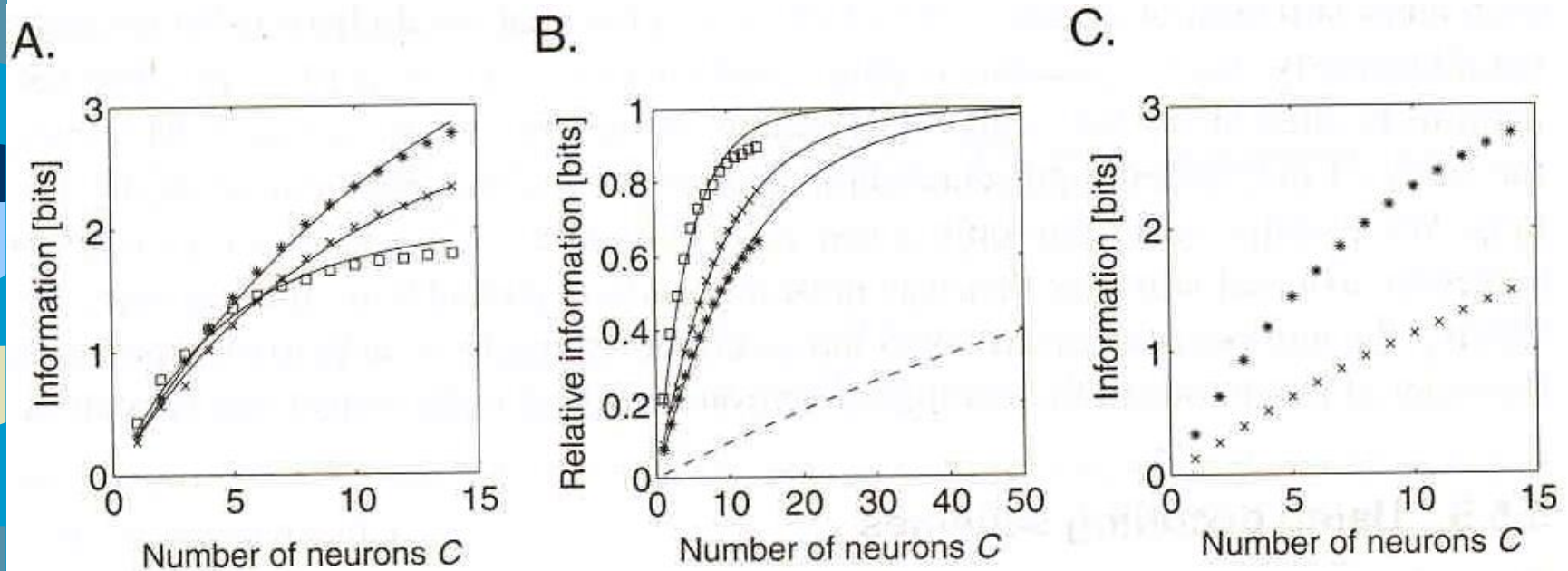


Fig. 5.15 (A) Estimate of mutual information between face stimuli and firing rate responses of C cells in the inferior-temporal cortex. The set of stimuli consisted of 20 faces (stars), 8 faces (crosses), and 4 faces (squares). (B) The information in the population of cells relative to the number of stimuli in the stimulus set. The solid lines are fits of the cell data to the ceiling model (eqn 5.51). The dashed line illustrates the values of the ceiling model for a stimulus set of 10,000 items and $y = 0.01$. (C) Estimate of mutual information with 20 faces when the neuronal response is derived from the spike count in 500 ms (stars) and 50 ms (crosses) [data from Rolls, Treves, and Tovee, *Exp. Brain Res.* 114: 149–62 (1997)].



Representación Distribuida de Información en el Cerebro

- Podemos observar que el máximo de las curvas depende del tamaño del estímulo. La curva se acerca más al comportamiento lineal cuando el conjunto de estímulos aumenta (20 caras).
- Los datos se pueden ajustar mediante la ecuación: $I(C) = S_{\max}(1 - (1 - y)^c)$, que es la línea continua sobre los símbolos.
- El parámetro “y” es la fracción de información llevada por cada neurona.
- S_{\max} es entropía máxima.



Representación Distribuida de Información en el Cerebro

- Todavía no hemos respondido a la pregunta de cómo de esparcida esta la información en la capa IT cortex.
- Para ello medimos el parámetro de esparcimiento $a = \langle r_i^s \rangle_{i,s}^2 / \langle (r_i^s)^2 \rangle_{i,s}$.
- En la siguiente figura se presenta la tasa de disparo de una neurona particular de IT cortex promediada sobre muchos experimentos y en respuesta a 70 estímulos individuales.
- Las respuestas son ordenadas de acuerdo que a los estímulos con las más grandes tasas de disparo se les da un número de estímulo más pequeño.



Representación Distribuida de Información en el Cerebro

- La línea punteada indica la tasa de disparo espontánea media de esta neurona, en relación a la cual consideramos la respuesta de la neurona.
- Vemos que las repuestas de las tasas de disparo son bastantes diferentes de la tasa de disparo espontánea.
- Aunque la mayoría de estas diferencias son menores.
- Si todas las diferencias fueran relevantes, diríamos que la neurona responde a todos los estímulos, por lo tanto el código que utiliza esta muy distribuido.
- Sin embargo, si solo las respuestas más diferenciabiles de la tasa de disparo espontánea son relevantes, concluiríamos que la neurona utiliza un código más esparcido ya que solo un pequeño número de estímulos conduce a la neurona a grandes respuestas.



Representación Distribuida de Información en el Cerebro

- Para tener en cuenta la relevancia en la tasa de disparo incluimos una *tasa de disparo umbral*.
- Así las tasas de disparo relevantes para una neurona dado un estímulo son aquellas que tienen una tasa de disparo media por encima de esa *tasa de disparo umbral* menos la tasa de disparo espontánea. Si están por debajo de ese umbral son cero y por tanto no se consideran relevantes.
- Estos valores que están por encima de esa tasa umbral son los que se utilizan para calcular el “esparcimiento”.
- El “esparcimiento”, $a = \langle r_i^s \rangle_{i,s}^2 / \langle (r_i^s)^2 \rangle_{i,s}$, se mide para 5 neuronas mostrándose la medida bastante baja.
- Solamente se tienen $a=0.3$ (grande) si consideramos umbrales muy pequeños, es decir desviaciones pequeñas del disparo espontáneo. El esparcimiento es más pequeño si tomamos diferencias más grandes con respecto a la espontánea para la tasa de disparo promedio.
- Esto nos da una idea del “esparcimiento” que hay en esta región cerebral que se ha analizado.

Representación Distribuida de Información en el Cerebro

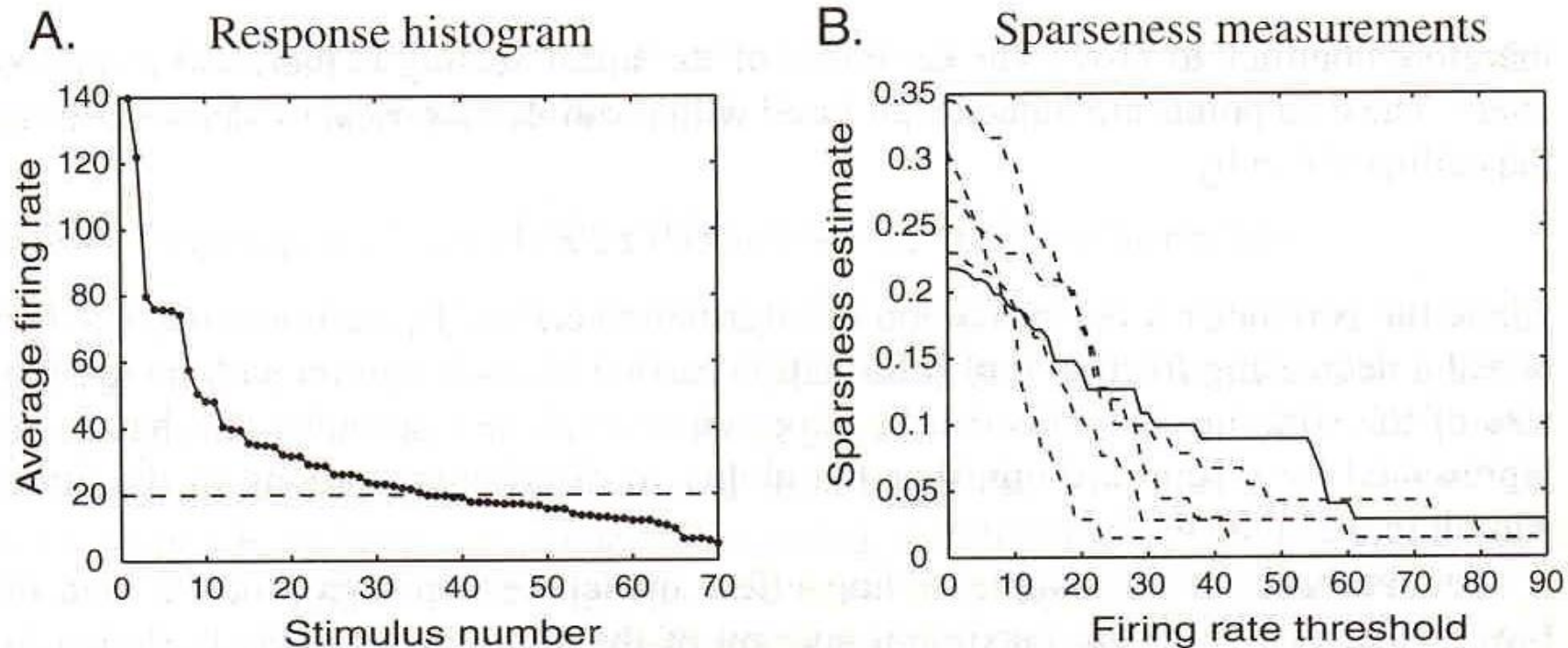


Fig. 5.16 (A) Average firing rates of a neuron in the IT cortex in response to each individual stimulus in a set of 70 stimuli (faces and objects). The responses are sorted in descending order. The horizontal dashed line indicates the spontaneous firing rate of this neuron. (B) The sparseness derived from five IT neurons calculated from eqn 5.47 in which responses lower than the firing rate threshold above the spontaneous firing rate are ignored. The solid line corresponds to the data shown in (A) [data from Rolls and Tovee, *J. Physiol.* 73: 713–26 (1995)].